Mutant deoxynucleotide carrier is associated with congenital microcephaly

Marjorie J. Rosenberg[1], Richa Agarwala[2], Gerard Bouffard[3], Joie Davis[1], Giuseppe Fiermonte[4], Mark S. Hilliard[5], Thorsten Koch[6], Linda M. Kalikin[7], Izabela Makalowska[1], D. Holmes Morton[8], Elizabeth M. Petty[7], James L. Weber[9], Ferdinando Palmieri[4], Richard I. Kelley[8,10], Alejandro A. Schäffer[2] & Leslie G. Biesecker[1]

The disorder Amish microcephaly (MCPHA) is characterized by severe congenital microcephaly, elevated levels of α-ketoglutarate in the urine and premature death[1]. The disorder is inherited in an autosomal recessive pattern and has been observed only in Old Order Amish families whose ancestors lived in Lancaster County, Pennsylvania. Here we show, by using a genealogy database and automated pedigree software, that 23 nuclear families affected with MCPHA are connected to a single ancestral couple. Through a whole-genome scan, fine mapping and haplotype analysis, we localized the gene affected in MCPHA to a region of 3 cM, or 2 Mb, on chromosome 17q25. We constructed a map of contiguous genomic clones spanning this region. One of the genes in this region, *SLC25A19*, which encodes a nuclear mitochondrial deoxynucleotide carrier (DNC)[2], contains a substitution that segregates with the disease in affected individuals and alters an amino acid that is highly conserved in similar proteins. Functional analysis shows that the mutant DNC protein lacks the normal transport activity, implying that failed deoxynucleotide transport across the inner mitochondrial membrane causes MCPHA. Our data indicate that mitochondrial deoxynucleotide transport may be essential for prenatal brain growth.

Microcephaly is caused by several environmental and genetic factors. It can occur as either part of a syndrome or isolation, and can occur with or without structural malformations of the brain[3]. The disorder is characterized by profound microcephaly, a congenital malformation of the brain and an inborn error of metabolism. At least 61 affected infants have been born in the past 40 years in 23 nuclear families. Affected individuals have head circumferences that are 6 to 12 standard deviations less the population mean and abnormally high concentrations of urinary α-ketoglutarate[1]. We considered that the increased levels of α-ketoglutarate might be the result of a defect in the α-ketoglutarate dehydrogenase complex, which comprises three enzymes: 2-oxoglutarate dehydrogenase (OMIM 203740), dihydrolipoamide succinyltransferase (OMIM 126063) and dihydrolipoamide dehydrogenase (OMIM 246900). We excluded genes encoding those enzymes on the basis of genetic linkage and haplotype analysis (data not shown), and concluded that this disorder must be associated with another genetic locus.
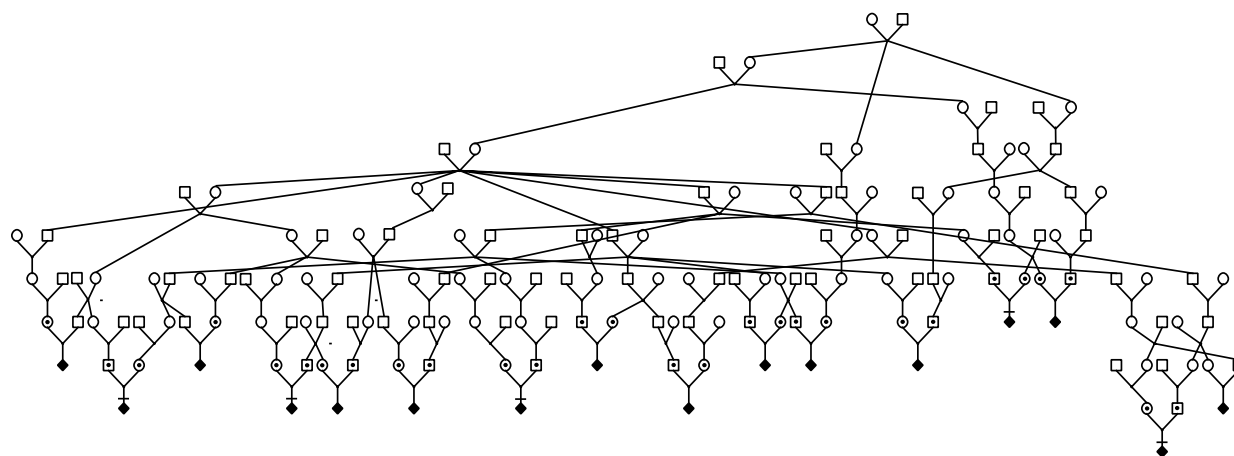

Fig. 1 Pedigree of the 16 nuclear affected families screened in this study. The pedigree was derived by searching a genealogy database and using a computational approach to determine the most recent common ancestor and the smallest overall pedigree for all 23 known nuclear families affected with MCPHA. Squares indicate males and circles indicate females. For clarity and privacy, each nuclear family is shown as having only one affected child of unspecified sex denoted by a solid diamond. Sampled carrier parents are denoted by an open symbol with a dot. One or more affected children were sampled from the families designated by a horizontal line above the diamond.

[1]National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, Bethesda, Maryland 20892-4472, USA. [2]National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland, USA. [3]NIH Intramural Sequencing Center, National Institutes of Health, Gaithersburg, Maryland, USA. [4]Department of Pharmaco-Biology, Laboratory of Biochemistry and Molecular Biology, University of Bari, Bari, Italy. [5]Logicon R.O.W. Sciences Corporation, Rockville, Maryland, USA. [6]Konrad-Zuse-Zentrum für Informationstechnik, Berlin, Germany. [7]University of Michigan Medical Center, Department of Internal Medicine, Ann Arbor, Michigan, USA. [8]The Clinic for Special Children, Strasburg, Pennsylvania, USA. [9]Marshfield Medical Research Foundation, Marshfield, Wisconsin, USA. [10]Kennedy-Krieger Institute and Department of Pediatrics, Johns Hopkins University, Baltimore, Maryland, USA. Correspondence should be addressed to M.J.R. (e-mail: marjr@nhgri.nih.gov).

We constructed a pedigree that connected the known nuclear families affected with MCPHA to a common ancestral couple. We ascertained a total of 23 nuclear families with members affected by MCPH, and included 16 of the families in the study (Fig. 1). We genotyped 20 individuals from eight families using a set of whole-genome markers[4] and calculated two-point log likelihood ratio (lod) scores for 365 autosomal markers assuming equal allele frequencies. Ten markers with either a lod score higher than 3.0, or a lod score higher than 2.0 and an adjacent marker with a lod score higher than 1.0, were identified. We analyzed these regions with additional markers and with samples that had been collected after the original whole-genome scan using two-point and multipoint linkage analyses (data not shown). After additional testing, only the markers on chromosome 17q24–25 had lod scores higher than 2.0.

We recalculated two-point lod scores using samples from 71 individuals (Methods and Table 1). We found eight consecutive markers from *D17S1301* to 5837.1 that were consistent with linkage in the sense that the peak two-point lod score was achieved at $\theta = 0.0$ and the peak two-point lod score was much higher than the standard threshold of 3.0. The lod scores differed substantially for some markers depending on whether we used equal allele frequencies or the allele frequencies specified by the Centre d'Etude du Polymorphisme Humain (CEPH) database, but among those markers, the lowest lod score at $\theta = 0$ was 5.10, which is significantly greater than the threshold for linkage.

Haplotype analysis showed that all 29 genotyped parents of an affected child shared a common haplotype that included ten microsatellite markers (Table 2*a*). This region extends from marker *D17S968* to marker 5837.1. We further narrowed down the critical region by recombination analysis in two unaffected children (Table 2*b*). An unaffected child in family 2 inherited the disease haplotype centromeric to and including *D17S1301* from both parents, whereas another unaffected child in family 15 inherited the disease haplotype telomeric to and including 5837.1 from both parents. Thus, we concluded that the locus associated with MCPHA lies between *D17S1301* and 5837.1.

To produce a physical map of this locus, we constructed a contiguous map of 79 BAC and PAC clones using a combination of bioinformatics and library screens with probes derived from genetic markers and BAC end sequences. Several maps of the telomeric portion of the candidate region for MCPHA were published during our study[5–7], and we used these as starting points for our map. The draft human sequence was also published during this work[8], and complemented or confirmed parts of the physical map described here. The contig of BAC and PAC clones extended about 3.7 Mb and contained 96 sequence-tagged sites (STSs), including 36 described STSs[5–7] and 60 markers identified in this study. Markers *D17S1301* and 5837.1 lie within the contig. Sequences of 21 of the clones from the complete map had been determined (3 completely and 18 partially), and 19 of them had been assembled into four sequence contigs that overlap with our map (Fig. 2). To fill in the gaps, we analyzed the sequences of three additional BACs contained in our candidate region. On the basis of the human genome working draft map (12 December 2000 freeze)[8], we estimated the length of the candidate region for MCPHA as roughly 2 Mb.

We identified a list of potential genes whose mutation might cause MCPHA. There were 53 full-length transcripts or known genes and 30 expressed-sequence tag (EST) clusters in the candidate region. We expanded this list after re-analyzing the assembled sequence using GeneMachine software[9] and BLAST analyses of the BACs whose sequences we had determined to connect the publicly available sequence information. From these analyses, we identified another 7 full-length cDNAs, 3 EST clusters and 11 predicted genes.

Because α-ketoglutarate is a component of the Krebs cycle, we reasoned that the α-ketoglutarate abnormality in MCPHA could be an indication of mitochondrial dysfunction. Two characterized genes from the list of candidates, *SLC25A19* (also known as *DNC* and *MUP1*) and *ATP5H*, encode proteins with known mitochondrial functions, and therefore we determined the sequences of these genes in individuals affected with MCPHA. The coding region of *ATP5H* was normal, but we identified a homozygous nucleotide change in the coding region of *DNC*, 530G→C (NM_021734), in an affected child. This mutation is predicted to prodcue a glycine-to-alanine amino-acid substitution at residue 177 (Gly177Ala). The mother of this affected child was heterozygous for the substitution. All affected children that we analyzed were homozygous for this alteration when tested with a PCR-amplified restriction-fragment length polymorphism generated by the mutation. We determined the prevalence of this substitution in individuals of non-Amish origin by the same technique. Eighty-six unrelated parents of individuals

Table 1 • Two-point lod scores of chromosome 17 markers

Marker	θ						Loops	Frequency*
	0	0.01	0.03	0.05	0.10	0.15		
D17S1351	4.56	4.43	4.19	3.94	3.28	2.63	14	CEPH
D17S1351	6.03	5.88	5.59	5.30	4.52	3.70	14	equal
D17S1862	−∞	15.81	16.12	15.81	14.31	12.40	16	CEPH
D17S1862	−∞	9.25	9.68	9.51	8.44	7.10	16	equal
D17S968	6.77	6.55	6.10	5.66	4.61	3.65	16	equal
16888.9	−∞	9.83	9.57	9.07	7.62	6.19	16	equal
D17S1301	13.14	12.73	11.93	11.14	9.23	7.45	16	CEPH
D17S1301	11.88	11.49	10.73	9.99	8.19	6.55	16	equal
D17S1839	5.10	4.92	4.56	4.22	3.43	2.72	16	CEPH
D17S1839	13.55	13.18	12.44	11.70	9.88	8.13	16	equal
D17S1603	5.40	5.17	4.73	4.31	3.35	2.54	16	CEPH
D17S1603	13.66	13.27	12.49	11.71	9.81	7.98	16	equal
15801.20A	9.42	9.08	8.42	7.78	6.28	4.93	16	equal
D17S785	15.70	15.25	14.37	13.49	11.33	9.27	16	CEPH
D17S785	14.27	13.85	13.01	12.19	10.18	8.28	16	equal
D17S1817	15.00	14.62	13.88	13.09	11.17	9.27	14	CEPH
D17S1817	9.81	9.50	8.87	8.25	6.77	5.40	14	equal
D17S801	14.40	14.01	13.23	12.45	10.53	8.67	16	CEPH
D17S801	11.97	11.62	10.92	10.23	8.54	6.95	16	equal
5837.1	12.05	11.70	11.00	10.32	8.64	7.07	16	equal
D17S722	5.97	8.63	8.71	8.40	7.24	5.95	15	equal
D17S939	−1.38	−0.23	0.68	1.14	1.59	1.60	16	CEPH
D17S939	−2.92	−1.66	−0.71	−0.19	0.43	0.61	16	equal
D17S802	−∞	−2.95	−1.87	−1.22	−0.28	0.16	13	CEPH
D17S802	−∞	−2.36	−1.28	−0.65	0.16	0.47	13	equal

*CEPH represents the allele frequency in the CEPH database; "equal" represents the allele frequency set to 1 divided by the number of alleles. θ, recombination fraction.

with undiagnosed multiple congenital anomaly syndrome[10] and forty unrelated individuals in the CEPH reference set (a total of 252 chromosomes analyzed) did not possess the substitution.

The *DNC* cDNA contains three mitochondrial carrier motifs that are characteristic of transport proteins found in the inner mitochondrial membrane[2,11]. The DNC protein is thought to facilitate the transport of deoxynucleotides from the cytosol into the mitochondrial matrix in exchange for ATP. Examination of the Conserved Domain Database[12] showed that the glycine residue that is altered in affected individuals is conserved in 272 of 275 sequences used to generate the consensus.

This suggests that this residue is highly conserved and that its mutation might disrupt function of the protein.

To test this hypothesis, we overexpressed wildtype human DNC and two mutants (Gly177Ala and Gly177Val) in *Escherichia coli*, purified them and reconstituted them into phospholipid vesicles. We then followed the time course of $[\alpha\text{-}^{35}\text{S}]$dATP/ADP exchange in proteoliposomes reconstituted with the recombinant wildtype DNC or the Gly177Ala mutant (Fig. 3). The wildtype protein efficiently catalyzed $[\alpha\text{-}^{35}\text{S}]$dATP/ADP exchange, whereas no transport activity was detected with the Gly177Ala mutant. The mutant Gly177Val was also inactive (data not

Table 2 • Haplotype analysis of families with *MCPHA*

Marker	D17S2193	D17S1351	D17S1862	D17S968	16888.9	D17S1301	22211.2	D17S1839	D17S1603	15801.20A	D17S785	D17S1817	D17S801	5837.1	D17S722	D17S939	D17S802
Position (cM)	89.3	96.0	97.6	NA	NA	100.0	NA	102.5	103.0	NA	103.5	103.5	103.5	NA	105.7	105.7	106.8
Parents[a]																	
2-1	153	177	225	170	182	159	1	245	226	183	208	207	284	201	152	207	184
2-2	150	179	213	170	182	159	1	245	226	183	208	207	284	201	152	207	184
3-1	144	179	213	0	182	159	0	245	226	183	208	207	284	201	0	207	184
3-2	0	179	213	170	182	159	0	245	226	183	208	207	284	201	152	207	184
4-1	144	179	213	170	182	159	0	245	226	183	208	207	284	201	152	205	182
4-2	141	179	213	170	182	159	0	245	226	183	208	207	284	201	152	207	182
5-1	0	179	213	170	182	159	0	245	226	183	208	207	0	201	0	205	0
5-2	141	177	213	170	182	159	0	245	226	0	208	207	284	201	0	205	182
6-1	0	179	213	170	0	159	0	245	226	0	208	207	284	201	152	0	180
6-2	144	179	213	0	182	159	0	245	226	183	208	207	0	201	0	0	180
7-2	141	179	213	170	182	159	0	245	226	183	208	207	284	201	152	203	182
8-1	144	179	213	170	182	159	1	245	226	183	208	207	284	201	152	207	182
8-2	144	179	213	170	182	159	0	245	226	183	208	207	284	201	152	203	182
9-1	144	179	213	170	182	159	0	245	226	183	208	207	284	201	152	209	180
9-2	153	177	213	170	182	159	0	245	226	183	208	207	284	201	152	211	180
10-1	153	179	205	170	0	159	0	245	226	0	208	0	284	201	152	0	182
10-2	156	179	213	170	182	159	0	245	226	183	208	0	284	201	152	207	182
11-1	144	179	213	170	182	159	0	245	226	0	208	0	284	201	0	209	182
11-2	141	179	213	170	182	159	0	245	226	183	208	0	284	201	152	207	182
12-2	144	179	213	170	182	159	0	245	226	183	208	207	284	201	152	211	182
13-1	0	0	213	170	182	159	0	245	226	183	208	207	284	201	152	207	0
13-2	0	0	213	170	182	159	0	245	226	183	208	207	284	201	152	207	0
14-1	0	0	213	170	182	159	0	245	226	183	208	207	284	201	152	207	0
14-2	0	0	213	170	182	159	0	245	226	183	208	207	284	201	152	201	0
15-1	0	0	213	170	182	159	0	245	226	183	208	207	284	201	152	209	0
15-2	0	0	213	170	182	159	0	245	226	183	208	207	284	201	152	209	0
16-1	0	0	213	170	182	159	0	245	226	183	208	207	284	201	158	203	0
16-2	0	0	213	170	182	159	0	245	226	183	208	207	284	201	152	209	0
17-1	0	0	207	170	182	159	0	245	226	183	208	0	284	201	0	0	0
Children[b]																	
2-9 pat	153	177	225	170	182	159	1	245	226	183	208	207	284	201	152	207	184
2-9 mat	150	179	213	170	182	159	2	247	240	179	212	205	308	201	179	203	178
15-7 pat	0	0	0	170	182	159	0	0	0	183	208	207	284	201	152	209	0
15-7 mat	0	0	0	170	170	163	0	0	0	179	212	207	264	201	152	209	0

[a]Haplotypes of the disease-linked chromosome for each screened parent are shown. The positions of the markers are taken from the Marshfield genetic map. For marker 22211.2, allele numbers are used. For all other markers, the allele size is given in base pairs. Samples with a '0' for a given marker were not typed. The shaded area represents the common haplotype shared by the carrier parent. The vertical lines mark the boundaries of the critical region as determined by the parental haplotypes. NA, not available. [b]Haplotypes of the two unaffected children that refined the candidate region associated with MCPHA. Both chromosomes are represented for each child[8]. The shaded area represents the portion of the disease-linked chromosomes that each child inherited from their parents. Child 2-9 inherited the disease chromosome from both parents from *D17S2193* to *D17S1301* and child 15-7 inherited the disease chromosome from both parents from 5837.1 to *D17S939*, and thus both of these regions are excluded. The refined critical region is marked by vertical lines.
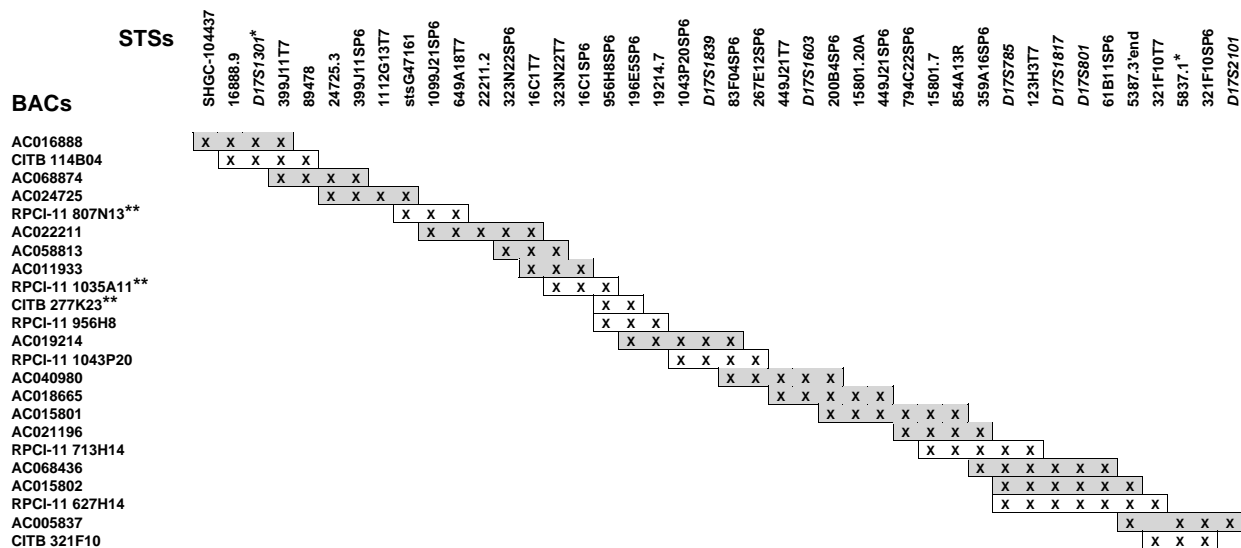
STSs

SHGC-104437, 16888.9, *D17S1301**, 399J11T7, 89478, 24725.3, 399J11SP6, 1112G13T7, stsG47161, 1099J21SP6, 649A18T7, 2221I.2, 323N22SP6, 16C1T7, 323N22T7, 16C1SP6, 956H8SP6, 196E5SP6, 19214.7, 1043P20SP6, *D17S1839*, 83F04SP6, 267E12SP6, 449J21T7, *D17S1603*, 200B4SP6, 15801.20A, 449J21SP6, 794C22SP6, 15801.7, 854A13R, 359A16SP6, *D17S785*, 123H3T7, *D17S1817*, *D17S801*, 61B11SP6, 5387.3end, 321F10T7, 5837.1*, 321F10SP6, *D17S2101*

BACs

AC016888
CITB 114B04
AC068874
AC024725
RPCI-11 807N13**
AC022211
AC058813
AC011933
RPCI-11 1035A11**
CITB 277K23**
RPCI-11 956H8
AC019214
RPCI-11 1043P20
AC040980
AC018665
AC015801
AC021196
RPCI-11 713H14
AC068436
AC015802
RPCI-11 627H14
AC005837
CITB 321F10

Fig. 2 Physical map of the candidate region of 2 Mb associated with MCPHA and bounded by the markers indicated in Table 2*b*. The candidate region is part of a larger contig of 3.7 Mb on chromosome 17q25. A minimum tiling path of BAC clones is shown, plus additional BACs for which sequence information exists. STS markers are shown on the horizontal axis from centromere to telomere. BAC clones are on the vertical axis. Accession numbers designate clones for which sequences are available in GenBank and are shaded gray. Other clones are denoted by their library and clone identification numbers. An asterisk indicates markers *D17S1301* and 5837.1, which define the critical region. Two asterisks indicate a BAC whose sequence was determined for this study. A full map containing all 79 BAC and PAC clones and 96 STS markers, and the primer sequences for the markers used in this analysis, are available from M.J.R. on request.

shown). Northern analysis and RT–PCR of samples from affected and unaffected individuals did not show differences in the levels of *DNC* RNA (data not shown).

We have used a positional cloning strategy to identify a gene that is mutated in individuals affected with MCPHA. Our data show that mitochondrial deoxynucleotide transport is crucial for the development of growth of the brain. We propose that insufficient transport of dNTPs into mitochondria in the developing central nervous system interferes with synthesis of mitochondrial DNA, causing abnormal brain growth. Our results indicate that other genes encoding proteins involved in dNTP metabolism or encoding purine or pyrimidine transporters may be candidate genes in the numerous other genetic loci associated with microcephaly.

Methods

Subjects. We collected basic clinical information on 23 nuclear families in the Old Order Amish population who had a child affected with MCPHA. Blood or buccal samples were collected from 71 individuals representing 16 of the 23 families, including 6 affected infants. The study was reviewed and approved by institutional review boards of the National Institutes of Health and Johns Hopkins University. Informed consent was obtained from, or on behalf of, all subjects.

Pedigree construction. Because manual genealogy analysis and pedigree drawing are prone to errors, we used a semiautomated database approach. The Amish Genealogy Database version 3.0 (ref. 13) includes information on nearly 300,000 individuals and was analyzed with the associated software PedHunter[14] to connect 23 nuclear families with an affected child into 'all-shortest-paths pedigrees' as described[15]. We constructed the pedigree four times by the same method as the number of known and screened families increased. The details of the latest construction are summarized here. Briefly, we used all 23 families to search for pedigrees that connected all parents of affected children to a common founder. We found ten distinct all-shortest-paths pedigrees and formulated a Steiner tree problem to extract a minimal pedigree (also called a Steiner pedigree) with the minimum number of meioses needed to provide one path of inheritance from the founder couple down to each carrier parent. We found an optimal solution to each of the ten Steiner pedigree problem examples using described software[16]. Among the solutions, we chose the one with the smallest number of individuals. For the final genetic linkage analyses, the subpedigree of the chosen Steiner pedigree was 'pruned' to remove seven unscreened families, leaving 16 screened nuclear families and their ancestors.

Fig. 3 Transport assays of wildtype and mutant DNC. Time course of [α-^{35}S]dATP/ADP exchange in proteoliposomes reconstituted with the recombinant wildtype (filled circles) or Gly177Ala mutant (open circles) DNC is shown. At time zero, 1 mM [α-^{35}S]dATP was added to proteoliposomes containing 10 mM ADP. The results shown are the mean of four experiments. No transport activity was detected with the mutant protein.

Genotyping. We tested several candidate markers in the vicinity of each of the three genes comprising the α-ketoglutarate complex by manual genotyping with a radioactive genotyping protocol[17]. We then carried out a
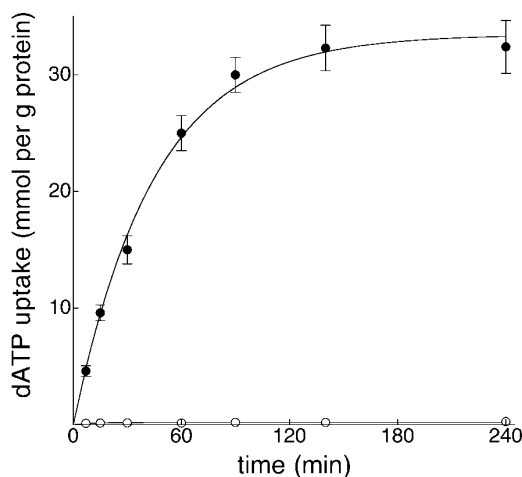
whole-genome screen with semiautomated genotyping[4] using the Marshfield screening set of nine markers (ResGen) on 20 subjects who were either parents or siblings of affected individuals or affected themselves, and who were selected according to the availability of large quantities of DNA (for example, those for whom blood samples were available). We isolated DNA from peripheral blood and buccal specimens by standard procedures. After preliminary linkage analysis, we tested additional markers flanking candidate loci using manual genotyping of these subjects and additional subjects collected after the whole-genome scan. To maintain and format the genotyping results, we designed a graphical user interface database application called FAST (families and samples tracker).

Linkage analysis. Linkage analyses were carried out by FASTLINK, version 4.1P (refs 18–20), and loops were broken as described[21]. In the initial genome scan, we used a smaller preliminary pedigree and carried out two-point analyses of each marker using equal allele frequencies. For the final linkage analysis, we used all loops in which one or two parents were genotyped and added loops with two non-genotyped parents when practical. In all analyses, we assumed that the locus associated with MCPHA was inherited in a fully penetrant autosomal recessive pattern and used a disease allele frequency of 0.035, corresponding to an incidence of about 1 in 800. Because it is difficult to estimate allele frequencies for genetic markers in closed populations, we carried out the linkage computations both with equal allele frequencies and with the allele frequencies specified by the CEPH database when available. For marker distances in multipoint calculations, we used the Marshfield map of chromosome 17 markers[22].

Genomic library screening and STS screening. We screened RPCI-11 BAC clones by filter hybridization using BAC end or STRP overgo probes[23]. We screened and analyzed other libraries by PCR amplification using STS primers and standard methods.

Construction of expression plasmids for human DNC mutants. The human *DNC* expression construct has been described[2]. We introduced two mutations (Gly177Ala and Gly177Val) into the wildtype human *DNC* cDNA by overlap-extension PCR[24]. Primers are available from F.P. on request. We expressed the wildtype and mutant proteins as inclusion bodies in the bacterial cytosol in *E. coli* BL21(DE3) as described[2,25]. We assayed the transport activities of the recombinant purified proteins as described[2,26]. The amount of the wildtype and DNC mutant proteins that was incorporated into liposomes varied between 18% and 25% of the protein added to the reconstitution mixture.

Databases. The FAST genotype database is available from mhilliar@nhgri.nih.gov. FASTLINK is available at ftp://fastlink.nih.gov/pub/fastlink. The Conserved Domain Database is available at http://www.ncbi.nlm.nih.gov:80/Structure/cdd/cdd.shtml.

Accession numbers. Sequence contigs overlapping MCPHA candidate region map: NT_024866, NT_010672, NT_010677 and NT_010641.

1. Kelley, R.I., Robinson, D., Puffenberger, E.G., Strauss, K.A. & Morton, D.H. Amish lethal microcephaly: a new metabolic disorder with severe congenital microcephaly and 2-ketoglutaric aciduria. *Am. J. Med. Genet.* (2002); advanced online publication, 7 June 2002 (doi:10.1002/ajmg 10529).
2. Dolce, V., Fiermonte, G., Runswick, M.J., Palmieri, F. & Walker, J.E. The human mitochondrial deoxynucleotide carrier and its role in the toxicity of nucleoside antivirals. *Proc. Natl Acad. Sci. USA* **98**, 2284–2288 (2001).
3. Carey, J.C. in *Rudolph's Pediatrics* (eds Rudolph, A.M., Hoffman, J.I.E. & Rudolph, C.D.) 427–429 (Appleton and Lange, Stamford, CT, 1996).
4. Weber, J. & Broman, K. Genotyping for human whole-genome scans: past, present, and future. *Adv. Genet.* **42**, 77–96 (2001).
5. Kuhlenbäumer, G. *et al*. A sequence-ready BAC/PAC contig and partial transcript map of approximately 1.5 Mb in human chromosome 17q25 comprising multiple disease genes. *Genomics* **62**, 242–250 (1999).
6. Kalikin, L.M. *et al*. An integrated physical and gene map of human distal chromosome 17q24-proximal 17q25 encompassing multiple disease loci. *Genomics* **57**, 36–42 (1999).
7. Risk, J.M. *et al*. The tylosis esophageal cancer (TOC) locus: more than just a familial cancer gene. *Dis. Esophagus* **12**, 173–176 (1999).
8. Lander, E.S. *et al*. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
9. Makalowska, I., Ryan, J.F. & Baxevanis, A.D. GeneMachine: gene prediction and sequence annotation. *Bioinformatics* **17**, 843–844 (2001).
10. Rosenberg, M.J. *et al*. Scanning for telomeric deletions and duplications and uniparental disomy using genetic markers in 120 children with malformations. *Hum. Genet.* **109**, 311–318 (2001).
11. Palmieri, F. Mitochondrial carrier proteins. *FEBS Lett.* **346**, 48–54 (1994).
12. Marchler-Bauer, A. *et al*. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30**, 281–283 (2002).
13. Agarwala, R., Schäffer, A.A. & Tomlin, J.F. Towards a complete North American Anabaptist Genealogy II: analysis of inbreeding. *Hum. Biol.* **73**, 533–545 (2001).
14. Agarwala, R., Biesecker, L.G., Hopkins, K.A., Francomano, C.A. & Schäffer, A.A. Software for constructing and verifying pedigrees within large genealogies and an application to the Old Order Amish of Lancaster County. *Genome Res.* **8**, 211–221 (1998).
15. Johnston, J.J. *et al*. A novel nemaline myopathy in the Amish caused by a mutation in troponin T1. *Am. J. Hum. Genet.* **67**, 814–821 (2000).
16. Koch, T. & Martin, A. Solving Steiner tree problems in graphs to optimality. *Networks* **32**, 207–232 (1998).
17. Biesecker, L.G. *et al*. Detection of a subtle rearrangement of chromosome 22 using molecular techniques. *Am. J. Med. Genet.* **58**, 389–394 (1995).
18. Cottingham, R.W. Jr, Idury, R.M. & Schäffer, A.A. Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* **53**, 252–263 (1993).
19. Lathrop, G.M., Lalouel, J.-M., Julier, C. & Ott, J. Strategies for multilocus analysis in humans. *Proc. Natl Acad. Sci. USA* **81**, 3443–3446 (1984).
20. Schäffer, A.A., Gupta, S.K., Shriram, K. & Cottingham, R.W. Jr. Avoiding recomputation in linkage analysis. *Hum. Hered.* **44**, 225–237 (1994).
21. Becker, A., Geiger, D. & Schäffer, A.A. Automatic selection of loop breakers for genetic linkage analysis. *Hum. Hered.* **48**, 49–60 (1998).
22. Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. & Weber, J.L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
23. Vollrath, D. in *Genome Analysis: A Laboratory Manual*, Vol. 4 (eds Birren, B. *et al*.) 187–216 (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1999).
24. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
25. Fiermonte, G., Walker, J.E. & Palmieri, F. Abundant bacterial expression and reconstitution of an intrinsic membrane-transport protein from bovine mitochondria. *Biochem. J.* **294**, 293–299 (1993).
26. Palmieri, F., Indiveri, C., Bisaccia, F. & Iacobazzi, V. Mitochondrial metabolite carrier proteins: purification, reconstitution, and transport studies. *Methods Enzymol.* **260**, 349–369 (1995).